SIAM J. NUMER. ANAL.
Vol. 47, No. 5, pp. 3464–3485

DISCRETIZATION OF THE FROBENIUS–PERRON OPERATOR USING A SPARSE HAAR TENSOR BASIS: THE SPARSE ULAM METHOD*

OLIVER JUNGE† AND PÉTER KOLTAI†

Abstract. The global macroscopic behavior of a dynamical system is encoded in the eigenfunctions of the associated Frobenius–Perron operator. For systems with low dimensional long term dynamics, efficient techniques exist for a numerical approximation of the most important eigenfunctions; cf. [M. Dellnitz and O. Junge, *SIAM J. Numer. Anal.*, 36 (1999), pp. 491–515]. They are based on a projection of the operator onto a space of piecewise constant functions supported on a neighborhood of the attractor—Ulam's method. In this paper we develop a numerical technique which makes Ulam's approach applicable to systems with higher dimensional long term dynamics. It is based on ideas for the treatment of higher dimensional partial differential equations using *sparse grids* [C. Zenger, *Sparse grids*, in Parallel Algorithms for Partial Differential Equations (Kiel, 1990), Vieweg, Braunschweig, 1991, pp. 241–251; H.-J. Bungartz and M. Griebel, *Acta Numer.*, 13 (2004), pp. 147–269]. Here, we use a sparse Haar tensor basis as the underlying approximation space. We develop the technique, establish statements about its complexity and convergence, and present two numerical examples.

Key words. Frobenius–Perron operator, transfer operator, Ulam's method, sparse grids

AMS subject classifications. 37M25, 65P99

DOI. 10.1137/080716864

1. Introduction.

Recently, numerical techniques have been developed which enable a coarse grained, yet global, statistical analysis of the long term behavior of certain dynamical systems. The basic algorithmic approach is to construct a box covering of some set of interest in phase space (e.g., the attractor of the system) [7, 8]. The cells in this covering then constitute the states of a finite Markov chain. The transition matrix of this chain (i.e., the matrix of transition probabilities between the boxes) can be viewed as a finite approximation to the Frobenius–Perron (or transfer) operator of the system. This operator describes how probability distributions on phase space evolve according to the dynamical system under consideration. In certain cases and in the appropriate functional analytic setting, eigenmodes of this operator can be used to characterize the long term behavior of the dynamics. Certain stationary distributions of the operator characterize how frequently typical trajectories visit certain parts of phase space. Eigenmodes at roots of unity enable the detection of macroscopic cycles in the dynamics, and eigenmodes at real eigenvalues close to one yield a decomposition of phase space into *almost invariant sets*, i.e., sets for which the probability for a typical point to be mapped back into the set is large [11]. The latter concept has been used, e.g., in order to detect and compute biomolecular conformations; cf. [13, 43, 44, 14].

Formally, the construction of the Markov chain can be viewed as projecting the transfer operator onto the space of functions which are piecewise constant on the elements of the box covering. Ulam conjectured [46] that for maps on the interval,

*Received by the editors February 27, 2008; accepted for publication (in revised form) July 14, 2009; published electronically October 29, 2009. The research of these authors was partly supported by the TopMath Ph.D. program within the Elite Network of Bavaria.
http://www.siam.org/journals/sinum/47-5/71686.html
†Center for Mathematics, Technische Universität München, D-85747 Garching, Germany (junge@ma.tum.de, koltai@ma.tum.de).

the stationary distribution of the chain converges to an invariant density of the map as the covering is refined. This has been proved for certain expanding maps by Li [39] and since then for various special classes of maps or stochastic processes also in higher dimensions [17, 31, 32, 16, 3, 20, 40, 41, 18, 29, 42, 21, 22, 30, 11, 33].

Ulam's method in combination with the subdivision approach from [7, 8] for the computation of the box covering works fine for systems with a low dimensional attractor; cf. also [5, 12]. For systems with higher dimensional long term dynamics the approach becomes inefficient due to the *curse of dimension*: the number of boxes in the covering scales exponentially in the dimension of the attractor. Adaptive approaches to the construction of the box covering [10, 34] do not remedy this fact.

In this paper we propose to attack this discretization task using ideas from *sparse grids* [45, 47, 4]. In this approach, which is being used, e.g., in the numerical solution of partial differential equations on higher dimensional domains, a basis of $[0,1]^d$ is built from a hierarchical basis of $[0,1]$ via a tensor product construction. The entire basis can be decomposed into subspaces which are spanned by basis functions of the same level of the one dimensional hierarchy in each factor. To each subspace one can associate its approximation *benefit* and its *cost* (which is typically given by its dimension). The idea of the sparse grid approach is to assemble a finite dimensional approximation space by choosing only those subspaces with the highest benefit to cost ratio.

In order to discretize the Frobenius–Perron operator, we employ a piecewise constant sparse hierarchical tensor basis (i.e., using the Haar system as the underlying one dimensional basis). This basis provides an approximation error of $\mathcal{O}(n^{-1}\cdot(\log\sqrt{n})^{d-1})$ for functions with bounded first derivatives, requiring a computational effort of $\mathcal{O}(n\cdot(\log\sqrt{n})^{d-1})$ (where n denotes the number of degrees of freedom in one coordinate direction and d is the dimension of phase space). In comparison, the standard Ulam basis requires $\mathcal{O}(n^d)$ basis functions in order to obtain an approximation error of $\mathcal{O}(n^{-1})$.

The paper is structured as follows: in section 2 we collect relevant basic concepts from dynamical systems theory, in particular Ulam's method. In section 3 we develop the sparse Ulam method by constructing the hierarchical tensor basis, deriving approximation properties, outlining the construction of the optimal approximation subspace, and comparing cost and accuracy of the new method with the standard Ulam approach. The section closes with statements about the convergence properties. In section 4 we collect considerations concerning an efficient implementation of our approach. In particular, we derive estimates on the computational effort as a function of the required accuracy. Section 5 presents two numerical examples: a comparison with Ulam's method for a three dimensional map with a smooth invariant density and a computation of the leading eigenfunctions of the transfer operator for a four dimensional map, constructed via a tensor product from two two dimensional standard maps.

Our implementation of the sparse Ulam method as well as the code for the example computations is freely available from the homepage of the authors.

2. The Frobenius–Perron operator and Ulam's method. We consider a discrete dynamical system given by a map

$$S : X \to X$$

on a compact subset X of $\mathbb{R}^d$. We assume S to be measurable with respect to the Borel-σ-algebra $\mathcal{B}$ on X.

2.1. The Frobenius–Perron operator. The *Frobenius–Perron operator* on Lebesgue integrable functions $f \in L^1 = L^1(X, \mathbb{C})$ is implicitly defined by

$$\int_A Pf \, \mathrm{d}m = \int_{S^{-1}(A)} f \, \mathrm{d}m \qquad \forall A \in \mathcal{B}; \tag{2.1}$$

cf. [38, 2]. The importance of this operator stems from the fact that certain eigenfunctions of it characterize the long term dynamical behavior of S in a statistical sense:

1. P is a Markov operator; assume it has the eigenvalue 1. An associated positive normalized eigenfunction f is the density of an *invariant measure* μ (i.e., this measure satisfies $\mu(A) = \mu(S^{-1}(A))$ for all $A \in \mathcal{B}$). According to the Birkhoff ergodic theorem [1], if μ is ergodic[1] and $\varphi : X \to \mathbb{R}$ is μ-integrable, then
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(S^k(x)) = \int_X \varphi \, \mathrm{d}\mu \tag{2.2}$$
for μ-almost all $x \in X$. For $\varphi = \chi_A$, the right-hand side counts the number of visits of the trajectory $(S^k(x))_{k\in\mathbb{N}}$ to the set A, and according to the theorem this just equals the μ-measure of A. If (as in our case) the ergodic measure is absolutely continuous, then (2.2) holds for a set of points with positive Lebesgue measure.
2. For certain maps one can prove that the Frobenius–Perron operator is quasi-compact as an operator on a suitable subspace of L^1 (for example, the space of functions of bounded variation [37]) and possesses isolated real eigenvalues close to 1 [6]. The associated eigenfunctions allow for a decomposition of state space into *almost invariant sets*, i.e., sets for which the invariance ratio
$$\alpha_m(A) := \frac{m(S^{-1}(A) \cap A)}{m(A)}$$
is close to 1 (m denotes Lebesgue measure) [11, 9, 25, 23, 24]. This idea has been used in order to classify and compute conformations of molecules [13, 44, 14, 15].
3. If roots of unity are eigenvalues of P, then the associated eigenfunctions can be used to detect (macroscopic) cycles in the dynamics of S [13].

2.2. Ulam's method. In order to approximate the (most important) eigenfunctions of the Frobenius–Perron operator, we have to discretize the corresponding infinite dimensional eigenproblem. To this end, we project the L^1 eigenvalue problem $Pf = \lambda f$ into a finite dimensional subspace. Let $V_n \subset L^1$ be an *approximation subspace* of L^1, and let $Q_n : L^1 \to V_n$ be some projection onto V_n. We then define the *discretized Frobenius–Perron operator* as

$$P_n := Q_n P \,|_{V_n}.$$

Ulam [46] proposed to use spaces of piecewise constant functions as approximation spaces: Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a disjoint partition of X. Define $V_n := \operatorname{span}\{\chi_1, \ldots, \chi_n\}$, where χ_i denotes the characteristic function of X_i. Further, let

$$Q_n h := \sum_{i=1}^{n} c_i \chi_i \qquad \text{with} \qquad c_i := \frac{1}{m(X_i)} \int_{X_i} h \, \mathrm{d}m,$$

[1] An invariant measure μ is ergodic if $\mu(A) \in \{0, 1\}$ for every invariant set $A \in \mathcal{B}$.

yielding $P_n\Delta_n^+ \subseteq \Delta_n^+$ and $P_n\Delta_n \subseteq \Delta_n$, where $\Delta_n := \left\{h \in V_n : \int |h|\, dm = 1\right\}$ and $\Delta_n^+ := \{h \in \Delta_n : h \geq 0\}$. Due to Brouwer's fixed point theorem, there always exists an approximate invariant density $f_n = P_n f_n \in \Delta_n^+$. The matrix representation of the linear map $P_n : \Delta_n \to \Delta_n$ with respect to the basis of characteristic functions is given by the *transition matrix* with entries

$$p_{ij} = \frac{m(X_j \cap S^{-1}(X_i))}{m(X_i)}. \tag{2.3}$$

Ulam conjectured [46] that if P has a unique stationary density f, then the sequence $(f_n)_{n\in\mathbb{N}}$ converges to f in L^1. It is still an open question under which conditions on S this is true in general. Li [39] proved the conjecture for expanding, piecewise continuous interval maps, and Ding and Zhou [18] proved it for the corresponding multidimensional case.

In [11], Ulam's method was applied to a small random perturbation of S which might be chosen such that the corresponding transfer operator is compact on L^2. In this case, perturbation results [36, section IV.3.5] for the spectrum of compact operators imply convergence.

Computing the transition matrix. The computation of one matrix entry (2.3) requires a d-dimensional quadrature. A standard approach to this is Monte Carlo quadrature (also cf. [28]), i.e.,

$$p_{ij} \approx \frac{1}{K}\sum_{k=1}^{K} \chi_i\left(S(x_k)\right), \tag{2.4}$$

where the points $x_1, \ldots, x_K$ are chosen independently and identically distributed (i.i.d.) from X_j according to a uniform distribution. In [27], a recursive exhaustion technique has been developed in order to compute the entries to a prescribed accuracy. However, this approach relies on the availability of local Lipschitz estimates on S which might not be cheaply computable in the case that S is given as the time-T-map of a differential equation.

For the Monte Carlo technique, consider a uniform partition of the unit cube into M^d congruent cubes of edge length $1/M$. Let P_M denote the transition matrix for this partition, and let $\tilde{P_M}$ be its Monte Carlo approximation. According to the central limit theorem (and its error estimate, the Berry–Esséen theorem [19]), we have[2]

$$|\tilde{p}_{ij} - p_{ij}| \lesssim 1/\sqrt{K} \tag{2.5}$$

for the absolute error of the entries of $\tilde{P}_M$. As a consequence, we need

$$\kappa \gtrsim \frac{M^d}{TOL^2} \tag{2.6}$$

sample points in total in order to achieve an absolute error of less than TOL for all the entries p_{ij}. Note that the accuracy of the entries of $\tilde{P}_M$ imposes a restriction on the achievable accuracy of the eigenvectors of P_M.

[2]We write $a(K) \lesssim b(K)$ if there is a constant $c > 0$ independent of K such that $a(K) \leq cb(K)$.

3. The sparse Ulam method. A naive application of Ulam's method to higher dimensional systems suffers from the *curse of dimension*: in order to achieve an L^1-accuracy of $\mathcal{O}(\varepsilon)$, one needs an approximation space of dimension $\mathcal{O}\left(\varepsilon^{-d}\right)$—translating into a prohibitively large computational effort for higher dimensional systems. There is a remedy to this problem for systems with low dimensional long term dynamics [8, 11]: the idea is to first compute a covering of the attractor of the system. On this (low dimensional) covering, Ulam's method can successfully be applied.

To avoid the exponential growth of complexity in the system (or attractor) dimension, we now follow an idea which was originally developed for quadrature problems [45] and used for the treatment of higher dimensional partial differential equations (cf., for example, [47, 4]): *sparse grids*. In fact, we change from the standard Ulam basis to a sparse hierarchical one in order to obtain a better cost/accuracy relation. In the following, we discuss the chosen basis in detail as well as its advantages and disadvantages.

3.1. The Haar basis. We describe the Haar basis on the d-dimensional unit cube $[0,1]^d$, deriving the multidimensional basis functions from the one dimensional ones; see, e.g., [26]. Let

$$f_{\mathrm{Haar}}(x) = -\mathrm{sign}(x) \cdot (|x| \leq 1), \tag{3.1}$$

where $(|x| \leq 1)$ equals 1 if the inequality is true and 0 otherwise. A basis function of the Haar basis is defined by the two parameters *level* i and *center (point)* j:

$$f_{i,j}(x) := \begin{cases} 1 & \text{if } i = 0, \\ 2^{\frac{i-1}{2}} \cdot f_{\mathrm{Haar}}\left(2^i\left(x - x_{i,j}\right)\right) & \text{if } i \geq 1, \end{cases} \tag{3.2}$$

where

$$x_{i,j} := (2j+1)/2^i, \qquad j \in \{0, \ldots, 2^{i-1} - 1\}. \tag{3.3}$$

A d-dimensional basis function is constructed from the one dimensional ones using a tensor product construction:

$$\varphi_{\ell,\mathbf{j}}(x) := \prod_{i=1}^{d} f_{\ell_i,\mathbf{j}_i}(x_i) \tag{3.4}$$

for $x = (x_1, \ldots, x_d) \in [0,1]^d$. Here $\ell = (\ell_1, \ldots, \ell_d)$, $\ell_i \in \{0, 1, 2, \ldots\}$, denotes the level of the basis function and $\mathbf{j} = (\mathbf{j}_1, \ldots, \mathbf{j}_d)$, $\mathbf{j}_i \in \{0, \ldots, 2^{\ell_i} - 1\}$, its center.

THEOREM 3.1 (Haar basis). *The set*

$$H = \left\{ f_{i,j} \mid i \in \mathbb{N}_0, j \in \{0, \ldots, 2^i - 1\} \right\}$$

is an orthonormal basis of $L^2([0,1])$, *the* Haar basis. *Similarly, the set*

$$H^d = \left\{ \varphi_{\ell,\boldsymbol{j}} \mid \ell \in \mathbb{N}_0^d, \boldsymbol{j}_i \in \{0, \ldots, 2^{\ell_i} - 1\} \right\}$$

is an orthonormal basis of $L^2([0,1]^d)$.

Figure 1 shows the basis functions of the first three levels of the one dimensional Haar basis. It will prove useful to collect all basis functions of one level in one subspace:

$$W_\ell := \mathrm{span}\left\{ \varphi_{\ell,\mathbf{j}} \mid \mathbf{j}_i \in \{0, \ldots, 2^{\ell_i} - 1\} \right\}, \qquad \ell \in \mathbb{N}_0^d. \tag{3.5}$$

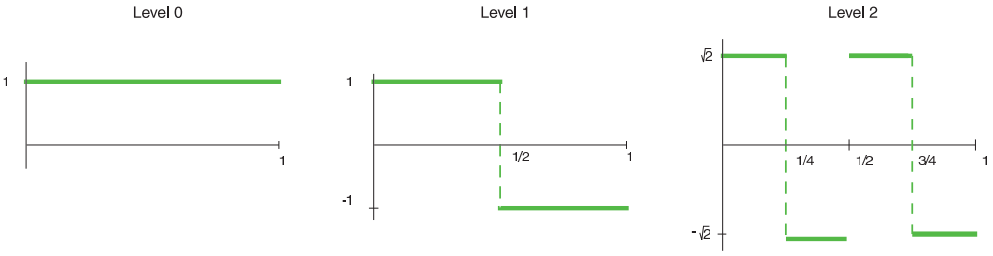

FIG. 1. *First three levels of the one dimensional Haar basis.*

Consequently, $L^2 = L^2([0,1]^d)$ can be written as the infinite direct sum of the subspaces W_ℓ,

$$L^2 = \bigoplus_{\ell \in \mathbb{N}_0^d} W_\ell. \tag{3.6}$$

In fact, it can also be shown that $L^1 = L^1([0,1]^d) = \bigoplus_{\ell \in \mathbb{N}_0^d} W_\ell$ as well. To see this, note that $\bigoplus_{\ell \in \mathbf{I}^d} W_\ell$ with $\mathbf{I} = \{\ell \mid \|\ell\|_\infty \leq n\}$ is the space of characteristic functions supported on the uniform decomposition of the unit cube in 2^n subcubes in every direction. Moreover, we have

$$\dim W_\ell = \prod_{i=1}^{d} 2^{\max\{0,\ell_i - 1\}} = 2^{\sum_{\ell_i \neq 0} \ell_i - 1}. \tag{3.7}$$

In order to get a finite dimensional approximation space most appropriate for our purposes, we are going to choose an optimal finite subset of the basis functions $\varphi_{\ell,\mathbf{j}}$. Since in general we do not have any a priori information about the function to be approximated, and since all basis functions in one subspace W_ℓ deliver the same contribution to the approximation error, we will use either all or none of them. In other words, the choice for the approximation space is transferred to the level of subspaces W_ℓ.

3.2. Approximation properties. The choice of the optimal set of subspaces W_ℓ relies on the contribution of each of these to the approximation error. The following statements give estimates of this.

LEMMA 3.2. *Let $f \in C^1([0,1])$, and let $c_{i,j}$ be its coefficients with respect to the Haar basis, i.e., $f = \sum_{ij} c_{i,j} f_{i,j}$. Then for $i > 0$ and all j*

$$|c_{i,j}| \leq 2^{-\frac{3i+1}{2}} \|f'\|_\infty.$$

For $f \in C^1\left([0,1]^d\right)$ we analogously have for $\ell \neq 0$ and all $\boldsymbol{j}$

$$|c_{\ell,\mathbf{j}}| \leq 2^{-\left(\sum_{\ell_i \neq 0} 3\ell_i + 1\right)/2} \prod_{\ell_i \neq 0} \|\partial_i f\|_\infty.$$

Proof. For $i \geq 1$

$$\begin{aligned} 2^{\frac{1-i}{2}} c_{ij} &= \int_{x_j - 2^{-i}}^{x_j} f - \int_{x_j}^{x_j + 2^{-i}} f \\ &= \int_{x_j - 2^{-i}}^{x_j} \left(f(x_j) + \int_{x_j}^{x} f' \right) \mathrm{d}x - \int_{x_j}^{x_j + 2^{-i}} \left(f(x_j) + \int_{x_j}^{x} f' \right) \mathrm{d}x, \end{aligned}$$

and thus

$$2^{\frac{1-i}{2}} |c_{ij}| \leq 2\|f'\|_\infty \int_0^{2^{-i}} x \, \mathrm{d}x,$$

which yields the claimed estimate for the one dimensional case. The bound in the d-dimensional case follows similarly. ∎

Using this bound on the contribution of a single basis function to the approximation of a given function f, we can derive a bound on the total contribution of a subspace W_ℓ. For $f_\ell \in W_\ell$

$$\|f_\ell\|_{L^1} \leq 2^{-\sum_{\ell_i \neq 0}(\ell_i+1)} \prod_{\ell_i \neq 0} \|\partial_i f\|_\infty, \tag{3.8}$$

$$\|f_\ell\|_{L^2} \leq 2^{-\sum_{\ell_i \neq 0}(\ell_i+3)/2} \prod_{\ell_i \neq 0} \|\partial_i f\|_\infty. \tag{3.9}$$

3.3. The optimal subspace. The main idea of the sparse grid approach is to choose *cost* and (approximation) *benefit* of the approximation subspace in an optimal way. We briefly sketch this idea here; for a detailed exposition see [47, 4]. For a set $\mathbf{I} \subset \mathbb{N}_0^d$ of multi-indices we define

$$W_\mathbf{I} = \bigoplus_{\ell \in \mathbf{I}} W_\ell.$$

Correspondingly, for $f \in L^1$, let $f_\mathbf{I} = \sum_{\ell \in \mathbf{I}} f_\ell$, where f_ℓ is the L^2-orthogonal projection[3] of f onto W_ℓ. We define the *cost* $C(\ell)$ of a subspace W_ℓ as its dimension,

$$C(\ell) = \dim W_\ell = 2^{\sum_{\ell_i \neq 0} \ell_i - 1}.$$

Since

$$\|f - f_\mathbf{I}\| \leq \sum_{\ell \notin \mathbf{I}} \|f_\ell\| = \sum_{\ell \in \mathbb{N}_0^d} \|f_\ell\| - \sum_{\ell \in \mathbf{I}} \|f_\ell\|, \tag{3.10}$$

the guaranteed increase in accuracy is bounded by the contribution of a subspace W_ℓ which we add to the approximation space. We therefore define the *benefit* $B(\ell)$ of W_ℓ as the upper bound on its L_1-contribution as derived above,

$$B(\ell) = 2^{-\sum_{\ell_i \neq 0}(\ell_i+1)}. \tag{3.11}$$

Note that we omitted the factor involving derivatives of f. The reason is that it does not affect the solution of the optimization problem (3.12).

Let $C(\mathbf{I}) = \sum_{\ell \in \mathbf{I}} C(\ell)$ and $B(\mathbf{I}) = \sum_{\ell \in \mathbf{I}} B(\ell)$ be the total cost and the total benefit of the approximation space $W_\mathbf{I}$. In order to find the optimal approximation space, we are now solving the following optimization problem: Given a bound $c > 0$ on the total cost, find an approximation space $W_\mathbf{I}$ which solves

$$\max_{C(\mathbf{I}) \leq c} B(\mathbf{I}). \tag{3.12}$$

[3]Note that since all functions in W_ℓ are piecewise constant and have compact support, this projection is well defined on L^1 as well.

One can show (cf. [4]) that $\mathbf{I} \subset \mathbb{N}_0^d$ is an optimal solution to (3.12) iff

$$\frac{C(\ell)}{B(\ell)} = \text{const} \qquad \text{for } \ell \in \partial\mathbf{I}, \tag{3.13}$$

where the *boundary* $\partial\mathbf{I}$ is given by $\partial\mathbf{I} = \{\ell \in \mathbf{I} \mid \ell' \in \mathbf{I}, \ell' \geq \ell \ \Rightarrow \ell' = \ell\}$.[4] Using the definitions for cost and benefit as introduced above, we obtain

$$\frac{C(\ell)}{B(\ell)} = \frac{2^{\sum_{\ell_i \neq 0}(\ell_i - 1)}}{2^{-\sum_{\ell_i \neq 0}(\ell_i + 1)}} = 2^{2\sum_{\ell_i \neq 0} \ell_i} = 2^{2|\ell|}, \tag{3.14}$$

where $|\ell|$ means the 1-norm of the vector ℓ.

The optimality condition (3.13) thus translates into the simple condition

$$|\ell| = \text{const} \quad \text{for } \ell \in \partial\mathbf{I}. \tag{3.15}$$

As a result, the optimal approximation space is $W_{\mathbf{I}(N)}$ with

$$\mathbf{I}(N) = \left\{\ell \in \mathbb{N}_0^d \mid |\ell| \leq N\right\}, \tag{3.16}$$

where the *level* $N = N(c) \in \mathbb{N}$ is dependent on the chosen cost bound c. Figure 2 schematically shows the basis functions of the optimal subspace in two dimensions for $N = 3$.

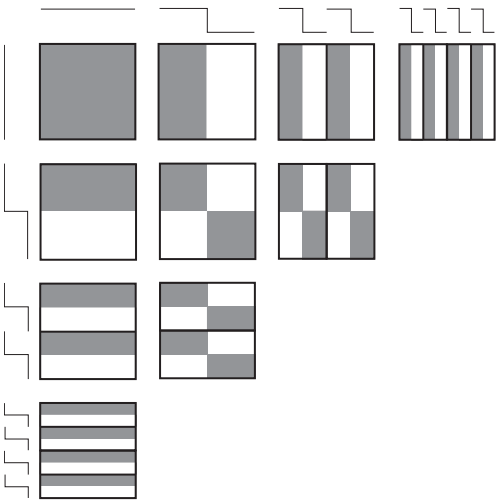

FIG. 2. *3rd level sparse basis in two dimensions. The shaded areas indicate the value* 1*, the white areas indicate the value* -1*, and the thicker lines are support boundaries.*

Remark 3.3. Because of the orthogonality of the Haar basis in L^2, one can take the squared contribution as the benefit in the L^2-case (resulting in equality in (3.10)). In this case we obtain the optimality condition

$$\sum_{\ell_i \neq 0} (\ell_i + 1) = \text{const} \quad \text{for } \ell \in \partial\mathbf{I} \tag{3.17}$$

[4] $\ell' \geq \ell$ is meant componentwise.

and correspondingly $W_{\mathbf{I}}$ with

$$\mathbf{I}(N) = \left\{ \ell \in \mathbb{N}_0^d : \sum_{\ell_i \neq 0} (\ell_i + 1) \leq N \right\}, \tag{3.18}$$

$N = N(c)$, as the optimal approximation space.

3.4. The discretized operator. Having chosen the optimal approximation space $V_N = W_{\mathbf{I}(N)}$, we now build the corresponding discretized Frobenius–Perron operator P_N. Since the sparse basis

$$B_N := \left\{ \varphi_{\ell,\mathbf{j}} \mid |\ell| \leq N, \mathbf{j}_i \in \{0, \dots, 2^{\ell_i} - 1\} \right\} \tag{3.19}$$

is an L^2-orthogonal basis of V_N, the natural projection $Q_N : L^2 \to V_N$ is given by

$$Q_N f = \sum_{\varphi \in B_N} \left(\int f\varphi \right) \varphi. \tag{3.20}$$

As noted in the previous section, the above definition of Q_N makes it well defined on L^1 as well. Choosing an arbitrary enumeration of the basis, the (*transition*) matrix of the discretized Frobenius–Perron operator

$$P_N = Q_N \circ P$$

with respect to B_N has entries

$$p_{ij} = \int \varphi_i \; P\varphi_j. \tag{3.21}$$

Writing $\varphi_i = \varphi_i^+ - \varphi_i^- = |\varphi_i| \cdot (\chi_i^+ - \chi_i^-)$, where $|\varphi_i|$ is the (constant) absolute value of the function over its support and χ_i^+ and χ_i^- are the characteristic functions on the supports of the positive and negative parts of φ_i, we obtain

$$p_{ij} = |\varphi_i||\varphi_j| \left(\int \chi_i^+ \; P\chi_j^+ - \int \chi_i^- \; P\chi_j^+ - \int \chi_i^+ \; P\chi_j^- + \int \chi_i^- \; P\chi_j^- \right), \tag{3.22}$$

which is, by (2.3),

$$p_{ij} = |\varphi_i||\varphi_j| \sum \pm m \left(X_j^\pm \cap S^{-1} \left(X_i^\pm \right) \right), \tag{3.23}$$

where $X_i^\pm = \operatorname{supp} \varphi_i^\pm$ and we add the four summands such as in (3.22). These can be computed in the same way as presented in section 2.

Remark 3.4. We note the following.

(a) If the ith basis function is the one corresponding to $\ell = (0, \dots, 0)$, then

$$p_{ij} = \delta_{ij}.$$

(b) The entries of P_N are bounded via

$$|p_{ij}| \leq \sqrt{\frac{m(X_j)}{m(X_i)}} \leq 2^{N/2}.$$

(c) If $P_N y = \lambda y$ for a $0 \neq y \in \mathbb{C}^{\dim V_N}$ with $\lambda \neq 1$, then $y_i = 0$ if the ith basis function is the one corresponding to $\ell = (0, \ldots, 0)$. This follows from

$$y_i \overset{(a)}{=} (e_i^\top P_N) y = e_i \lambda y = \lambda y_i. \tag{3.24}$$

It is straightforward to show that this property is shared by every Ulam-type projection method with a constant function as an element of the basis of the approximation space. This observation is useful for the reliable computation of an eigenvector at an eigenvalue close to one (since it is badly conditioned): (3.24) allows us to reduce the eigenproblem to the subspace orthogonal to the constant function.

With the given change in (c), properties (a)–(c) are valid for the numerical realization as well.

3.5. Convergence. As has been pointed out in the introduction and in section 2.2, statements about the convergence of Ulam's method exist in certain cases. For certain random perturbations of S we obtain the convergence of the sparse Ulam method by applying the same arguments as for Ulam's method [11] and the following lemma. An open question is if, in general, the convergence of Ulam's method implies convergence of sparse Ulam and vice versa.

LEMMA 3.5. $\|Q_N f - f\|_{L^p} \overset{n\to\infty}{\longrightarrow} 0$ *for* $f \in L^\infty$, $p = 1, 2$.

Proof. The convergence in the L^2-norm is trivial.

Since X is bounded, we have $L^\infty(X) \subset L^2(X) \subset L^1(X)$. Moreover, there is a constant $c_2 > 0$ such that

$$\|h\|_{L^1} \leq c_2 \|h\|_{L^2} \qquad \forall h \in L^2(X).$$

Thus convergence in the L^2-norm also implies the convergence in the L^1-norm. ∎

4. Complexity. In this section, we collect basic statements about the complexity of both methods.

4.1. Cost and accuracy. We defined the total cost of an approximation space as its dimension and the accuracy via its *contribution* or *benefit*; see (3.11). In this section we derive a recurrence formula for these numbers, depending on the *level* of the optimal subspaces and the system dimension.

Let $C(N, d)$ be the dimension of $W_{\mathbf{I}(N)}$ in phase space dimension d. Then

$$C(N, d) = C(N, d-1) + \sum_{k=1}^{N} C(N-k, d-1) 2^{k-1}, \tag{4.1}$$

since if $\ell = (*, \ldots, *, 0)$, then the last dimension plays no role in the number of basis functions, and the total number of basis functions for such ℓ's is $C(N, d-1)$. If, on the other hand, $\ell = (*, \ldots, *, \ell_d)$ with $\ell_d > 0$, then the number of basis functions with such ℓ's is $C(N - \ell_d, d-1) 2^{\ell_d - 1}$, because there are $2^{\ell_d - 1}$ one dimensional basis functions of level ℓ_d possible for the tensor product in the last dimension. For $d = 1$ we simply deal with the standard Haar basis, so $C(N, 1) = 2^N$.

LEMMA 4.1.

$$C(N, d) \doteq \frac{N^{d-1}\ 2^{N-d+1}}{(d-1)!}, \tag{4.2}$$

where $\doteq$ *means the leading order term in* N.

Proof. The proof is by induction on d. The claim holds clearly for $d = 1$. Assume that it holds for $d-1$. By considering the recurrence formula (4.1), we see that $C(N,d) = p(N)\ 2^N$, where p is a polynomial of order less than or equal to d. Consequently,

$$\begin{aligned} C(N,d) &\doteq \frac{N^{d-2}\ 2^{N-d+2}}{(d-2)!} + \sum_{k=1}^{N} \frac{(N-k)^{d-2}\ 2^{N-k-d+2}}{(d-2)!} 2^{k-1} \\ &= \frac{N^{d-2}\ 2^{N-d+2}}{(d-2)!} + \frac{2^{N-d+1}}{(d-2)!} \sum_{k=1}^{N} (N-k)^{d-2} \\ &\doteq \frac{N^{d-2}\ 2^{N-d+2}}{(d-2)!} + \frac{2^{N-d+1}}{(d-2)!} \frac{N^{d-1}}{d-1} \\ &\doteq \frac{N^{d-1}\ 2^{N-d+1}}{(d-1)!}. \qquad \square \end{aligned}$$

According to (3.10), the approximation error $\|f - f_{\mathbf{I}}\|$ is bounded by $\sum_{\ell \notin \mathbf{I}} \|f_\ell\|$, i.e.,

$$\|f - f_{\mathbf{I}}\| \leq \sum_{|\ell|>N} \|f_\ell\|,$$

if we use the optimal approximation space $W_{\mathbf{I}(N)}$. By (3.8) this means

$$\|f - f_{\mathbf{I}}\| \leq \sum_{|\ell|>N} \left[2^{-\sum_{\ell_i \neq 0}(\ell_i+1)} \prod_{\ell_i \neq 0} \|\partial_i f\|_\infty \right].$$

Again, the constants $\prod_{\ell_i \neq 0} \|\partial_i f\|_\infty$ depend only on the function to be approximated. Thus, without a priori knowledge about f, we need to assume that they can be bounded by some common constant and accordingly define the discretization error of the Nth level sparse basis as

$$E(N,d) = \sum_{|\ell|>N} 2^{-\sum_{\ell_i \neq 0}(\ell_i+1)}. \tag{4.3}$$

Let $E(-n,d)$ for $n \in \mathbb{N}, n > 0$ represent the error of the empty basis and $\ell = (\tilde{\ell}, \ell_d)$ with $\tilde{\ell} \in \mathbb{N}_0^{d-1}$. Then

$$\begin{aligned} E(N,d) &= \sum_{|\ell|>N} 2^{-\sum_{\ell_i \neq 0}(\ell_i+1)} \\ &= \sum_{\ell_d=0}^{\infty} 2^{-(\ell_d+1)(\ell_d \neq 0)} \sum_{|\tilde{\ell}|>N-\ell_d} 2^{-\sum_{\tilde{\ell}_i \neq 0}(\tilde{\ell}_i+1)} \\ &= \sum_{\ell_d=0}^{\infty} 2^{-(\ell_d+1)(\ell_d \neq 0)} E(N-\ell_d, d-1), \end{aligned}$$

where the expression $(\ell_i \neq 0)$ has the value 1 if it is true and the value 0 otherwise. This leads, by splitting the sum, to the recurrence formula

$$E(N,d) = E(N,d-1) + \sum_{k=1}^{N} E(N-k,d-1) 2^{-k-1} + \underbrace{\sum_{k=N+1}^{\infty} 2^{-k-1} E(-1,d-1)}_{=2^{-N-1}E(-1,d-1)}. \tag{4.4}$$

We easily compute that $E(N,1) = 2^{-N-1}$ for $N \geq 0$ and $E(-1,d) = (3/2)^d$.

LEMMA 4.2.

$$E(N,d) \doteq \frac{N^{d-1}\ 2^{-N-d}}{(d-1)!}, \tag{4.5}$$

where, again, $\doteq$ *means the leading order term in* N.

Proof. The proof is by induction on d. The claim holds for $d = 1$; assume that it holds for $d-1$. Then

$$\begin{aligned} E(N,d) &\doteq \frac{N^{d-2}2^{-N-d+1}}{(d-2)!} + \sum_{k=1}^{N} \frac{(N-k)^{d-2}2^{-N+k-d+1}}{(d-2)!} 2^{-k-1} + \left(\frac{3}{2}\right)^{d-1} 2^{-N-1} \\ &\doteq \frac{N^{d-2}2^{-N-d+1}}{(d-2)!} + \frac{2^{-N-d}}{(d-2)!} \sum_{k=1}^{N} (N-k)^{d-2} \\ &\doteq \frac{2^{-N-d}}{(d-2)!} \frac{N^{d-1}}{d-1}. \end{aligned}$$

∎

Comparison with Ulam's method. We now compare the expressions for the asymptotic behavior of cost and discretization error dependent on the discretization level N and the problem dimension d in Lemmas 4.1 and 4.2 to the corresponding expressions for the standard Ulam basis, i.e., the span of the characteristic functions on a uniform partition of the unit cube into cubes of edge length 2^{-M} in each coordinate direction—this is $\bigoplus_{\|\ell\|_\infty \leq M} W_\ell$. This space consists of $(2^M)^d$ basis functions, and the discretization error is $\mathcal{O}\left(2^{-M}\right)$.

We thus have—up to constants—the following asymptotic expressions for cost and error of the sparse and the standard bases:

	Cost	Error
Sparse basis	$(N/2)^{d-1}\,2^N$	$(N/2)^{d-1}\,2^{-N}$
Standard basis	2^{dM}	2^{-M}

To highlight the main difference, consider the following simple computation: The expressions for the errors are equal if

$$M = N + d - (d-1)\log_2 N.$$

Using this value for M in the cost expression, we get $N^{d-1}\ 2^{N-d} < 2^{dN+d^2-d(d-1)\log_2 N}$, i.e.,

$$\frac{N}{d+1} > \log_2 N - 1, \tag{4.6}$$

as a sufficient condition for the sparse basis to be more efficient than the standard basis. Since we neglected constants and lower order terms in this estimate, the only conclusion we can draw from this is that from a certain accuracy requirement on, the sparse basis is more efficient than the standard one.

4.2. Computing the matrix entries. When we use Monte Carlo quadrature in order to approximate the entries of the transition matrix in both methods, the overall computation breaks down into the following three steps:

1. mapping the sample points,
2. constructing the transition matrix, and

3. solving the eigenproblem.

While steps 1 and 3 are identical for both methods, step 2 differs significantly. This is due to the fact that, in contrast to Ulam's method, the basis functions of the sparse hierarchical tensor basis have global and nondisjoint supports.

4.2.1. Number of sample points. Applying Monte Carlo approximation to (3.22), we obtain

$$\tilde{p}_{ij} = |\varphi_i||\varphi_j| \left(\frac{m\left(X_j^+\right)}{K_j^+} \sum_{k=1}^{K_j^+} \chi_i^+\left(S(x_k^+)\right) - \chi_i^-\left(S(x_k^+)\right) - \frac{m\left(X_j^-\right)}{K_j^-} \sum_{k=1}^{K_j^-} \chi_i^+(S(x_k^-)) - \chi_i^-(S(x_k^-)) \right), \tag{4.7}$$

where the sample points $x_k^\pm$ are chosen i.i.d. from a uniform distribution on X_j^+ and X_j^-, respectively. In fact, since the union of the supports of the basis functions in one subspace W_ℓ covers all of X, we can reuse the same set of κ sample points and their images for each of the subspaces W_ℓ (i.e., $\binom{N+d}{d}$ times). Note that the number $K_j^\pm$ of test points chosen in $X_j^\pm$ now varies with j since the supports of the various basis functions are of different sizes: on average, $K_j^\pm = \kappa m(X_j^\pm)$. Accordingly, for the absolute error of $\tilde{p}_{ij}$ we get

$$|\tilde{p}_{ij} - p_{ij}| \sim \frac{m(X_j)}{\sqrt{m(X_i)m(X_j)}\sqrt{\kappa\, m(X_j)}} = \frac{1}{\sqrt{\kappa m(X_i)}}, \tag{4.8}$$

where we used that $m(X_i^\pm) \sim m(X_i)$. In the worst case we thus get $|\tilde{p}_{ij} - p_{ij}| \sim 2^{N/2}/\sqrt{\kappa}$, which implies

$$\kappa \gtrsim \frac{2^N}{TOL^2} \tag{4.9}$$

for the total number of test points required in order to achieve an accuracy of TOL in the entries of the transition matrix.

Comparison with Ulam's method. Aiming at a final accuracy of $\varepsilon > 0$ of the eigenvector, we have to choose M and N accordingly. Assuming that the corresponding eigenproblems are well conditioned, $TOL = \varepsilon$ is a reasonable choice for the required accuracy of the entries. This implies a number of

$$\kappa \gtrsim \varepsilon^{-(d+2)} \tag{4.10}$$

sample points for the standard realization of Ulam's method (cf. section 2.2) and yields, since $2^N \lesssim \varepsilon^{-1}\left(\log(\varepsilon^{-1})\right)^{d-1}$,

$$\kappa \gtrsim \varepsilon^{-3}\left(\log(\varepsilon^{-1})\right)^{d-1} \tag{4.11}$$

sample points for the sparse Ulam method. Note that for $d \geq 2$, the sparse Ulam method requires fewer sample points than Ulam's method in order to achieve comparable accuracy in the eigenvector approximation. In Figure 3 we compare the estimates (2.6) and (4.9) graphically for $TOL = \varepsilon = 0.01$.

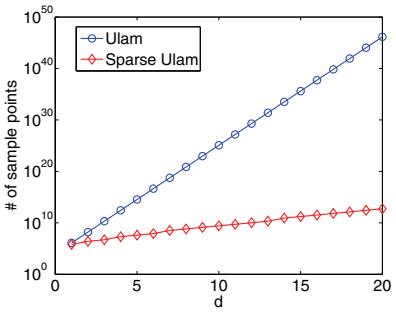

FIG. 3. *Estimated number of sample points dependent on the dimension of phase space.*

4.2.2. Number of index computations. While in Ulam's method each sample point is used in the computation of one entry of the transition matrix only, this is not the case in the sparse Ulam method. In fact, each sample point (and its image) is used in the computation of $|\mathbf{I}(N)|^2$ matrix entries, namely, one entry for each pair (W_ℓ, W_m) of subspaces.

Correspondingly, for each sample point x (and its image) and for each $\ell \in \mathbf{I}(N)$, we have to compute the index $\mathbf{j}$ of the basis function $\varphi_{\ell,\mathbf{j}} \in W_\ell$ whose support contains x. Since (cf. the previous section) the required number of sample points is $\mathcal{O}(\frac{2^N}{TOL^2})$ and $|\mathbf{I}(N)| = \binom{N+d}{d}$, this leads to

$$\kappa \cdot |\mathbf{I}(N)| = \frac{2^N}{TOL^2}\binom{N+d}{d} \lesssim \frac{1}{TOL^2}\frac{N^d}{d!}2^N \doteq \frac{2^{d-1}N}{d}\frac{1}{TOL^2}\dim V_N$$

of these computations (for reasonable d). In contrast, in Ulam's method, the corresponding number is

$$\kappa \cdot 1 = \frac{2^{dM}}{TOL^2} = \frac{1}{TOL^2}\dim V_M.$$

Note that for the sparse Ulam method the number of index computations is not staying proportional to the dimension of the approximation space. However, it is still scaling much more mildly with d than for Ulam's method. Figure 4 shows a graphical comparison of the estimated number of index computations dependent on the dimension of the problem for a prescribed error tolerance $TOL = 0.01$.

4.2.3. Occupancy of the transition matrix. The matrix which represents the discretized transfer operator in Ulam's method is *sparse*: the supports of the basis functions are disjoint, and thus $p_{ij} \neq 0$ only if $S(X_j) \cap X_i \neq \emptyset$. Hence, for a sufficiently fine partition, the number of partition elements X_i which are intersected by the image $S(X_j)$ is determined by the local expansion of S. This is a fixed number related to a Lipschitz estimate on S, and so the matrix of the discretized transfer operator with respect to the standard Ulam basis is sparse for sufficiently large n. Unfortunately this property is not shared by the matrix with respect to the sparse basis, as the following considerations show.

The main reason for this is that the supports of the basis functions in the sparse basis are not localized; cf. the thin and long supports of the basis of W_ℓ for $\ell = (N, 0, \ldots, 0)$. This means that the occupancy of the transition matrix strongly

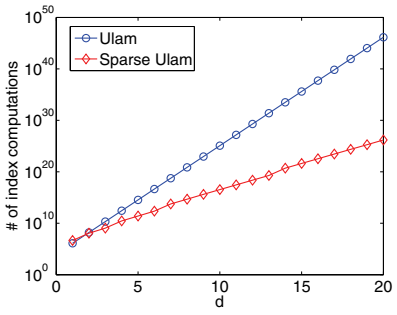

FIG. 4. *Estimated number of index computations dependent on the dimension of phase space.*

depends on the *global behavior* of the dynamical system S. Let

$$B_\ell := \{\varphi_{\ell,\mathbf{j}} \mid \mathbf{j}_i \in \{0,\dots,2^{\ell_i}-1\}\}$$

denote the basis of W_ℓ, and let

$$\mathrm{nnz}(\mathbf{k},\ell) = |\{(i,j) \mid S(\mathrm{supp}(\varphi_i)) \cap \mathrm{supp}(\varphi_j) \neq \emptyset, \varphi_i \in B_{\mathbf{k}}, \varphi_j \in B_\ell\}|$$

be the number of nonzero matrix entries which arise from the interaction of the basis functions from the subspaces $W_{\mathbf{k}}$ and W_ℓ if $W_{\mathbf{k}}$ is mapped. We define the *matrix occupancy* of a basis $B_{\mathbf{I}} = \bigcup_{\ell\in\mathbf{I}} B_\ell$ as

$$\mathrm{nnz}(B_{\mathbf{I}}) = \sum_{\mathbf{k},\ell\in\mathbf{I}} \mathrm{nnz}(\mathbf{k},\ell). \tag{4.12}$$

In order to estimate $\mathrm{nnz}(\mathbf{k},\ell)$, we employ upper bounds L_i, $i = 1,\dots,d$, for the Lipschitz constants of S; cf. Figure 5. We obtain the following proposition.

PROPOSITION 4.3.

$$\mathrm{nnz}(\boldsymbol{k},\ell) \leq |B_{\boldsymbol{k}}| \prod_{i=1}^{d} \left\lceil \frac{L_i \cdot 2^{-\boldsymbol{k}_i+1-(\boldsymbol{k}_i=0)}}{2^{-\ell_i+1-(\ell_i=0)}} \right\rceil. \tag{4.13}$$

Proof. Since we have used upper bounds for the Lipschitz constants, *one* mapped box has at most the extension $L_i \cdot 2^{-\mathbf{k}_i+1-(\mathbf{k}_i=0)}$ in the ith dimension. Consequently, its support intersects with at most

$$\left\lceil \frac{L_i \cdot 2^{-\mathbf{k}_i+1-(\mathbf{k}_i=0)}}{2^{-\ell_i+1-(\ell_i=0)}} \right\rceil$$

supports of basis functions from W_ℓ. ∎

Remark 4.4. Numerical experiments suggest that the above bound approximates the matrix occupancy pretty well. However, it could be improved: (3.21) shows that a matrix entry could still be zero even if $\mathrm{supp}(\varphi_i)$ and $\mathrm{supp}(P\varphi_j)$ intersect. This is the case, e.g., if $\mathrm{supp}(P\varphi_j)$ is included in a subset of $\mathrm{supp}(\varphi_i)$, where φ_i is constant (i.e., does not change sign). The property $\|Pf\|_{L^1} = \|f\|_{L^1}$ for $f \geq 0$ and positivity (see [38]) of P imply $p_{ij} = 0$, since $\left\|\varphi_j^+\right\|_{L^1} = \left\|\varphi_j^-\right\|_{L^1}$.

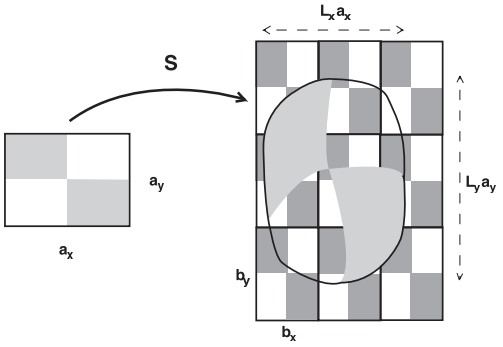

FIG. 5. *Model for the matrix occupancy in two dimensions. Shaded and colorless (white) parts show the function values ($\pm|\varphi|$), thicker black lines the support boundaries.*

An asymptotic estimate. Let us examine $\mathrm{nnz}(\mathbf{k}, \ell)$ for $\mathbf{k} = (0, \ldots, 0, N)$ and $\ell = (N, 0, \ldots, 0)$. By taking all Lipschitz constants $L_i = 1$, we get

$$\mathrm{nnz}(\mathbf{k}, \ell) \gtrsim 2^{2N},$$

since $|B_{\mathbf{k}}| = 2^{N-1}$ and the image of each basis function from $B_{\mathbf{k}}$ intersects with each basis function from B_ℓ. Since $|B_N| \approx N^{d-1}2^N$, we get

$$2^{2N} \lesssim \mathrm{nnz}(B_N) \lesssim N^{2d-2}2^{2N}. \tag{4.14}$$

The exponential term dominates the polynomial one for large N, so asymptotically we will not get a sparse matrix.

Does this affect the calculations regarding efficiency made above? As already mentioned, the error of Ulam's method is $\varepsilon = \mathcal{O}(2^{-M})$ while its cost is $2^{dM} = \mathcal{O}(\varepsilon^{-d})$. Assuming that the sparse Ulam method has the same error $\varepsilon = \mathcal{O}(N^{d-1}2^{-N})$, its worst-case cost is

$$\mathcal{O}(N^{2d-2}2^{2N}) \lesssim \varepsilon^{-2}N^{4d-4} \lesssim \varepsilon^{-2}\log(\varepsilon^{-1})^{4d-4},$$

where we used $N^{d-1}2^{-N} \lesssim 2^{-N/2}$, which leads to $\log(\varepsilon) \lesssim -N$. Clearly, this means—similarly to subsection 4.1—partially overcoming the curse of dimension. Even in the most optimistic case, i.e., the costs are $\mathcal{O}(2^{2N})$, we have at least $\mathcal{O}(\varepsilon^{-2}N^{2d-2})$ costs, so the sparse Ulam method is more efficient than Ulam's (concerning the number of flops for a matrix-vector-multiplication) only if $d \geq 3$.

However, the fact that the transition matrix with respect to the sparse basis is not sparse poses another obstacle: the memory requirements for storing the matrix grow faster with the dimension d of phase space than one might expect from the above estimates about the complexity of the sparse Ulam method. Figure 6 shows a comparison of the estimated number of nonzero entries dependent on d. Clearly, for $d > 4$ the storage requirements render computations on standard workstations impossible.

5. Numerical examples.

5.1. A three dimensional expanding map. We compare both methods by approximating the invariant density of a simple three dimensional map. Let S_i :

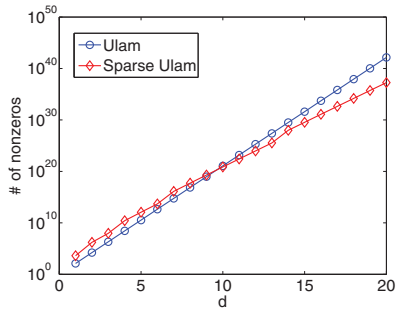

FIG. 6. *Estimated number of nonzero entries in the matrix representation of the discretized operator dependent on the dimension of phase space for* $\varepsilon = 0.01$.

$[0,1] \to [0,1]$ be given by

$$S_1(x) = 1 - 2|x - 1/2|,$$
$$S_2(x) = \begin{cases} 2x/(1-x), & x < 1/3, \\ (1-x)/(2x), & \text{else}, \end{cases}$$
$$S_3(x) = \begin{cases} 2x/(1-x^2), & x < \sqrt{2} - 1, \\ (1-x^2)/(2x), & \text{else}, \end{cases}$$

and $S : [0,1]^3 \to [0,1]^3$ be the tensor product map $S(x) = (S_1(x_1), S_2(x_2), S_3(x_3))^\top$, where $x = (x_1, x_2, x_3)^\top$. This map is expanding, and its unique invariant density is given by (cf. [18])

$$h(x) = \frac{8}{\pi(1 + x_3^2)(1 + x_2)^2}.$$

We approximate h by Ulam's method on an equipartition of 2^{3M} boxes for $M = 4, 5, 6$ as well as by the sparse Ulam method on levels $N = 4, 5, 6$.

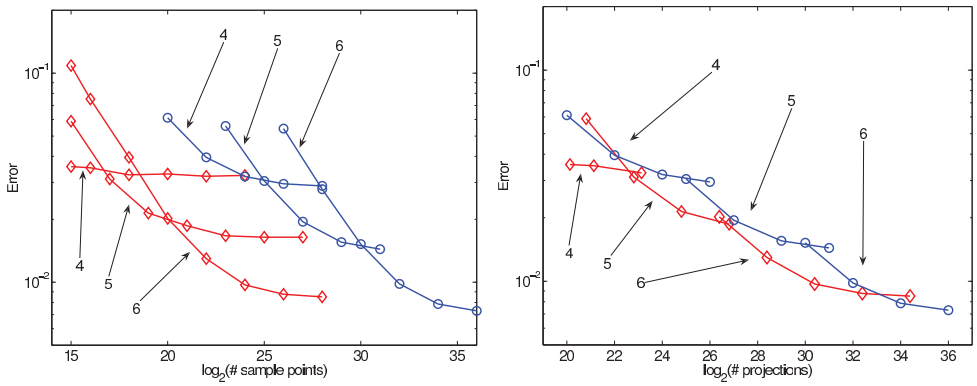

FIG. 7. *Left:* L^1*-error of the approximate invariant density dependent on the number of sample points for levels* $N, M = 4, 5, 6$. *Right: Corresponding number of index computations.*

Figure 7 shows the L^1-error for both methods dependent on the number of sample points (left) as well as the number of index computations along these curves (right).

While the sparse Ulam method requires almost three orders of magnitude fewer sample points than Ulam's method, the number of index computations is roughly comparable. This is in good agreement with our theoretical considerations in sections 4.2.1 and 4.2.2.

In Figure 8 we show the dependence of the L^1-error on the number of floating point operations for levels $M, N = 3, \ldots, 6$. Again, the sparse Ulam method is ahead of Ulam's method by almost an order of magnitude.

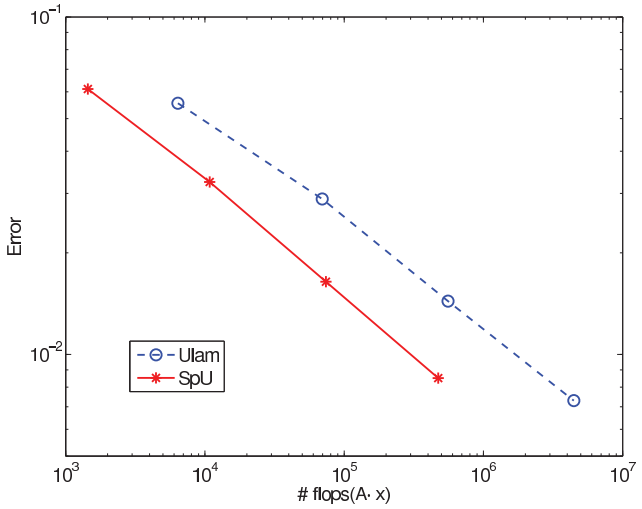

FIG. 8. *L^1-error of the approximate invariant densities dependent on the number of nonzeros in the transition matrices.*

5.2. A four dimensional conservative map. In a second numerical experiment, we approximate a few dominant eigenfunctions of the transfer operator for an area-preserving map. Since the information on almost invariant sets does not change [23] (but the eigenproblem becomes easier to solve), we here consider the symmetrized transition matrix $\frac{1}{2}(P + P^\top)$; cf. also [35].

Consider the so-called *standard map* $S_\rho : [0,1]^2 \to [0,1]^2$,

$$(x_1, x_2)^\top \mapsto (x_1 + x_2 + \rho \sin(2\pi x_1) + 0.5, x_2 + \rho \sin(2\pi x_1))^\top \mod 1,$$

where $0 < \rho < 1$ is a parameter. This map is *area-preserving*; i.e., the Lebesgue measure is invariant with respect to S_ρ. Figure 9 shows approximations of the eigenfunctions at the second largest eigenvalue of S_ρ for $\rho = 0.3$ (left) and $\rho = 0.6$ (right) computed via Ulam's method on an equipartition of $2^{2\cdot 6}$ boxes (i.e., for $M = 6$).

We now define $S : [0,1]^4 \to [0,1]^4$ by

$$S = S_{\rho_1} \otimes S_{\rho_2},$$

with $\rho_1 = 0.3$ and $\rho_2 = 0.6$. Note that the eigenfunctions of S are tensor products of the eigenfunctions of the S_{ρ_i}. This is reflected in Figures 10 and 11, where we show the eigenfunctions at the two largest eigenvalues, computed by the sparse Ulam method on level $N = 8$, using 2^{24} sample points overall. Clearly, each of these two is a tensor product of the (two dimensional) eigenfunction at the second largest eigenvalue with the (two dimemsional) invariant (i.e., constant) density.

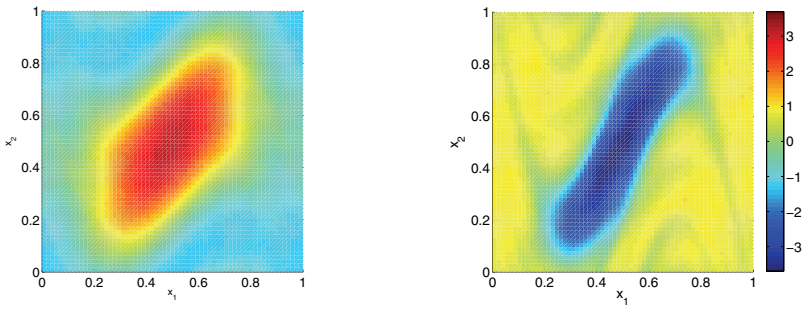

FIG. 9. *Eigenfunction of the symmetrized transition matrix at the second largest eigenvalue for the standard map. Left:* $\rho = 0.3$, $\lambda_2 = 0.97$. *Right:* $\rho = 0.6$, $\lambda_2 = 0.93$.

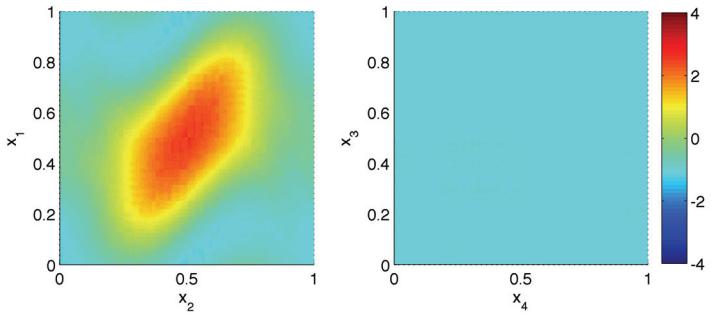

FIG. 10. *Approximate eigenfunction at* $\lambda_2 = 0.97$. *Left:* $v_2(\cdot, \cdot, x_3, x_4)$ *for fixed* x_3, x_4. *Right:* $v_2(x_1, x_2, \cdot, \cdot)$ *for fixed* x_1, x_2.

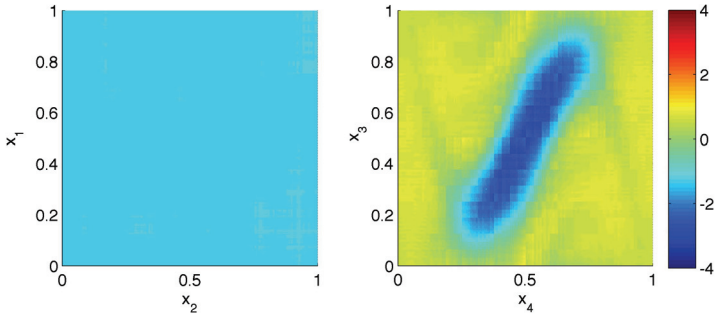

FIG. 11. *Approximate eigenfunction at* $\lambda_2 = 0.93$.

Figure 12 shows an eigenfunction for which both factors of the tensor product are nonconstant. The resolution of this eigenfunction seems worse than for those with one constant factor. In fact, for an approximation of an eigenfunction which is constant with respect to, say, x_3 and x_4, it suffices to consider subspaces W_ℓ with $\ell = (\ell_1, \ell_2, 0, 0)$. All other coefficients are zero, the problem reduces to a two dimensional one, and so the eigenfunctions are not perturbed by basis functions varying in the x_3 and x_4 directions.

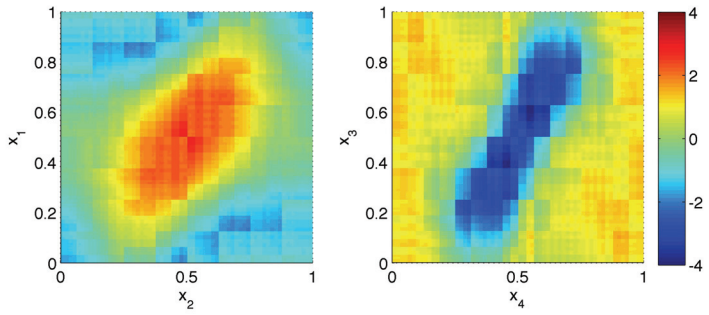

FIG. 12. *Approximate eigenfunction at* $\lambda = 0.80$.

6. Conclusion and outlook. The sparse Ulam method offers a promising way for the numerical analysis of the Frobenius–Perron operator in systems with high dimensional long term dynamics. The computational complexity scales much more mildly with the dimension of phase space (resp., the dimension of the invariant set one is interested in) than the standard approach.

One important open question is how to deal with the fact that the transition matrix resulting from the sparse Ulam discretization is not sparse. It might be possible to exploit the fact that matrix entries associated to coarser subspaces can be computed from those on the finest level on the fly. Another idea is to suitably modify the cost measure C associated to each subspace. This will be the subject of future investigations.

REFERENCES

[1] G. D. BIRKHOFF, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA, 17 (1931), pp. 650–660.
[2] A. BOYARSKY AND P. GÓRA, *Laws of Chaos*, Probability and Its Applications, Birkhäuser Boston, Boston, 1997.
[3] A. BOYARSKY, P. GÓRA, AND Y.-S. LOU, *Constructive approximations to the invariant densities of higher-dimensional transformations*, Constr. Approx., 10 (1994), pp. 1–13.
[4] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numer., 13 (2004), pp. 147–269.
[5] M. DELLNITZ, G. FROYLAND, AND O. JUNGE, *The algorithms behind GAIO–set oriented numerical methods for dynamical systems*, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, B. Fiedler, ed., Springer, Berlin, 2001, pp. 145–174, 805–807.
[6] M. DELLNITZ, G. FROYLAND, AND S. SERTL, *On the isolated spectrum of the Perron-Frobenius operator*, Nonlinearity, 13 (2000), pp. 1171–1188.
[7] M. DELLNITZ AND A. HOHMANN, *The computation of unstable manifolds using subdivision and continuation*, in Nonlinear Dynamical Systems and Chaos, H. Broer, S. van Gils, I. Hoveijn, and F. Takens, eds., Progr. Nonlinear Differential Equations Appl. 19, Birkhäuser, Basel, 1996, pp. 449–459.
[8] M. DELLNITZ AND A. HOHMANN, *A subdivision algorithm for the computation of unstable manifolds and global attractors*, Numer. Math., 75 (1997), pp. 293–317.
[9] M. DELLNITZ AND O. JUNGE, *Almost invariant sets in Chua's circuit*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 7 (1997), pp. 2475–2485.
[10] M. DELLNITZ AND O. JUNGE, *An adaptive subdivision technique for the approximation of attractors and invariant measures*, Comput. Vis. Sci., 1 (1998), pp. 63–68.
[11] M. DELLNITZ AND O. JUNGE, *On the approximation of complicated dynamical behavior*, SIAM J. Numer. Anal., 36 (1999), pp. 491–515.
[12] M. DELLNITZ AND O. JUNGE, *Set oriented numerical methods for dynamical systems*, in Handbook of Dynamical Systems, Vol. 2, B. Fiedler, ed., North–Holland, Amsterdam, 2002, pp. 221–264.

[13] P. Deuflhard, M. Dellnitz, O. Junge, and C. Schütte, *Computation of essential molecular dynamics by subdivision techniques*, in Computational Molecular Dynamics: Challenges, Methods, Ideas, P. Deuflhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R. Skeel, eds., Lect. Notes Comput. Sci. Eng. 4 , Springer, Berlin, 1999, pp. 98–115.
[14] P. Deuflhard and C. Schütte, *Molecular conformation dynamics and computational drug design*, in Applied Mathematics Entering the 21st Century, J. M. Hill and R. Moore, eds., SIAM, Philadelphia, 2004, pp. 91–119.
[15] P. Deuflhard and M. Weber, *Robust Perron cluster analysis in conformation dynamics*, Linear Algebra Appl., 398 (2005), pp. 161–184.
[16] J. Ding, Q. Du, and T. Y. Li, *High order approximation of the Frobenius-Perron operator*, Appl. Math. Comput., 53 (1993), pp. 151–171.
[17] J. Ding and T.-Y. Li, *Markov finite approximation of the Frobenius-Perron operator*, Nonlinear Anal., 17 (1991), pp. 759–772.
[18] J. Ding and A. Zhou, *Finite approximations of Frobenius-Perron operators. A solution of Ulam's conjecture on multi-dimensional transformations*, Phys. D, 92 (1996), pp. 61–68.
[19] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol.* 2., 2nd ed., Wiley, New York, 1971.
[20] G. Froyland, *Finite approximation of Sinai-Bowen-Ruelle measures for Anosov systems in two dimensions.*, Random Comput. Dynam., 3 (1995), pp. 251–263.
[21] G. Froyland, *Approximating physical invariant measures of mixing dynamical systems in higher dimensions*, Nonlinear Anal., 32 (1998), pp. 831–860.
[22] G. Froyland, *Ulam's method for random interval maps*, Nonlinearity, 12 (1999), pp. 1029–1052.
[23] G. Froyland, *Statistically optimal almost-invariant sets*, Phys. D, 200 (2005), pp. 205–219.
[24] G. Froyland, *Unwrapping eigenfunctions to discover the geometry of almost-invariant sets in hyperbolic maps*, Phys. D, 237 (2008), pp. 840–853.
[25] G. Froyland and M. Dellnitz, *Detecting and locating near-optimal almost-invariant sets and cycles*, SIAM J. Sci. Comput., 24 (2003), pp. 1839–1863.
[26] M. Griebel, P. Oswald, and T. Schiekofer, *Sparse grids for boundary integral equations*, Numer. Math., 83 (1999), pp. 279–312.
[27] R. Guder, M. Dellnitz, and E. Kreuzer, *An adaptive method for the approximation of the generalized cell mapping*, Chaos Solitons Fractals, 8 (1997), pp. 525–534.
[28] F. Y. Hunt, *A Monte Carlo approach to the approximation of invariant measures*, Random Comput. Dynam., 2 (1994), pp. 111–133.
[29] F. Y. Hunt, *Approximating the invariant measures of randomly perturbed dissipative maps*, J. Math. Anal. Appl., 198 (1996), pp. 534–551.
[30] F. Y. Hunt, *Unique ergodicity and the approximation of attractors and their invariant measures using Ulam's method*, Nonlinearity, 11 (1998), pp. 307–317.
[31] F. Y. Hunt and W. M. Miller, *On the approximation of invariant measures*, J. Statist. Phys., 66 (1992), pp. 535–548.
[32] F. Y. Hunt and W. M. Miller, *On the approximation of invariant measures*, J. Statist. Phys., 66 (1992), pp. 535–548.
[33] M. S. Islam, P. Góra, and A. Boyarsky, *Approximation of absolutely continuous invariant measures for Markov switching position dependent random maps*, Int. J. Pure Appl. Math., 25 (2005), pp. 51–78.
[34] O. Junge, *An adaptive subdivision technique for the approximation of attractors and invariant measures: Proof of convergence*, Dyn. Syst., 16 (2001), pp. 213–222.
[35] O. Junge, J. Marsden, and I. Mezic, *Uncertainty in the dynamics of conservative maps*, in Proceedings of the 43rd IEEE Conference on Decision and Control, Vol. 2, IEEE Control Systems Society, Piscataway, NJ, 2004, pp. 2225–2230.
[36] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verag, New York, 1984.
[37] G. Keller, *On the rate of convergence to equilibrium in one-dimensional systems*, Comm. Math. Phys., 96 (1984), pp. 181–193.
[38] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise*, 2nd ed., Appl. Math. Sci. 97, Springer-Verlag, New York, 1994.
[39] T.-Y. Li, *Finite approximation for the Frobenius–Perron operator. A solution to Ulam's conjecture.*, J. Approx. Theory, 17 (1976), pp. 177–186.
[40] W. M. Miller, *Stability and approximation of invariant measures for a class of nonexpanding transformations*, Nonlinear Anal., 23 (1994), pp. 1013–1025.
[41] W. M. Miller, *Frobenius–Perron operators and approximation of invariant measures for set-valued dynamical systems*, Set-Valued Anal., 3 (1995), pp. 181–194.

[42] W. M. Miller, *Discrete approximation of invariant measures for multidimensional maps*, in African Americans in Mathematics (Piscataway, NJ, 1996), DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 34, AMS, Providence, RI, 1997, pp. 29–46.
[43] C. Schütte, *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*, Habilitation thesis, Free University Berlin, Berlin, Germany, 1999.
[44] C. Schütte, W. Huisinga, and P. Deuflhard, *Transfer operator approach to conformational dynamics in biomolecular systems*, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, B. Fieder, ed., Springer, New York, 2001, pp. 191–223.
[45] S. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 148 (1963), pp. 1042–1045.
[46] S. M. Ulam, *A Collection of Mathematical Problems*, Interscience, New York, 1960.
[47] C. Zenger, *Sparse grids*, in Parallel Algorithms for Partial Differential Equations (Kiel, 1990), Notes Numer. Fluid Mech. 31, Vieweg, Braunschweig, 1991, pp. 241–251.